Dashboard Visual Analytics para Análisis de Tráfico de Red CICIDS2017

- Princce Yorwin Mariños Hilario

Los ataques cibernéticos afectan la seguridad de redes y sistemas críticos.
Es necesario analizar grandes volúmenes de datos de tráfico para detectar patrones maliciosos.
Esto afecta ya que que para volúmenes de datos sumamente grandes como lo es el trafico de red, esto hace que se pierdan en los datos las ip maliciosas y esto provoca que no se pueda obtener la información requerida de ella.

Lo que se propone

- Permitir la identificación de patrones y características distintivas entre tráfico benigno y malicioso.
- Apoyar la toma de decisiones en ciberseguridad mediante visualizaciones intuitivas.

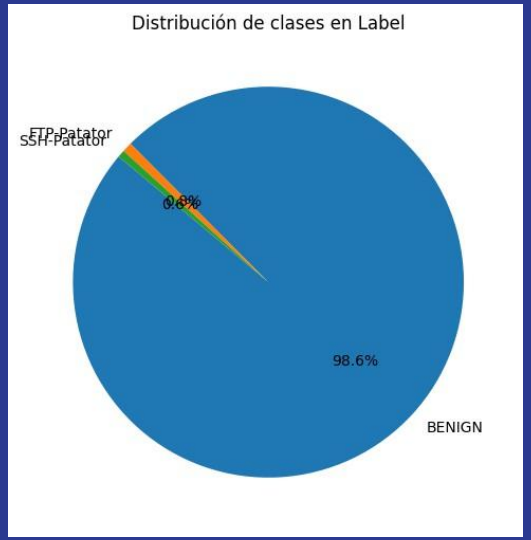
Dataset CICIDS2017

- El dataset CICIDS2017 contiene datos de tráfico de red con más de 80 variables para analizar tráfico benigno y ataques (DoS, PortScan, BruteForce, etc.).
- Variables clave usadas en el análisis:
  - **Flow Duration**: duración del flujo de conexión (importante para medir tiempo activo).
  - **Fwd IAT Total** y **Bwd IAT Total**: tiempos entre paquetes enviados hacia adelante y hacia atrás (clave para distinguir tipos de tráfico).
  - **Label**: clase que indica si el flujo es benigno o malicioso.
  - **Timestamp**: marca temporal para análisis de tendencias y comportamientos temporales.
- Este contexto ayuda a entender cómo se comporta el tráfico y qué variables son críticas para detectar ataques.

	count	mean	std	min	25%	50%	75%	max
Flow Duration	975530.0	10570777.51	29126736.72	0.0	181.0	31301.0	393419.5	119999987.0
Fwd IAT Total	975562.0	10283969.87	28973325.06	0.0	1.0	4.0	267366.0	120000000.0
Bwd IAT Total	975562.0	9449795.25	28121199.85	0.0	0.0	3.0	60869.5	120000000.0

¿Qué Problemas identifican en el DataSet CICIDS2017?

- Desbalance extremo en la variable **Label**: 98% benigno, menos del 2% ataques.
- Presencia de valores negativos erróneos en **Flow Duration**, corregidos para evitar distorsión.
- Muchas variables con alta proporción de ceros o datos faltantes (flags TCP, métricas bulk), descartadas por falta de información.

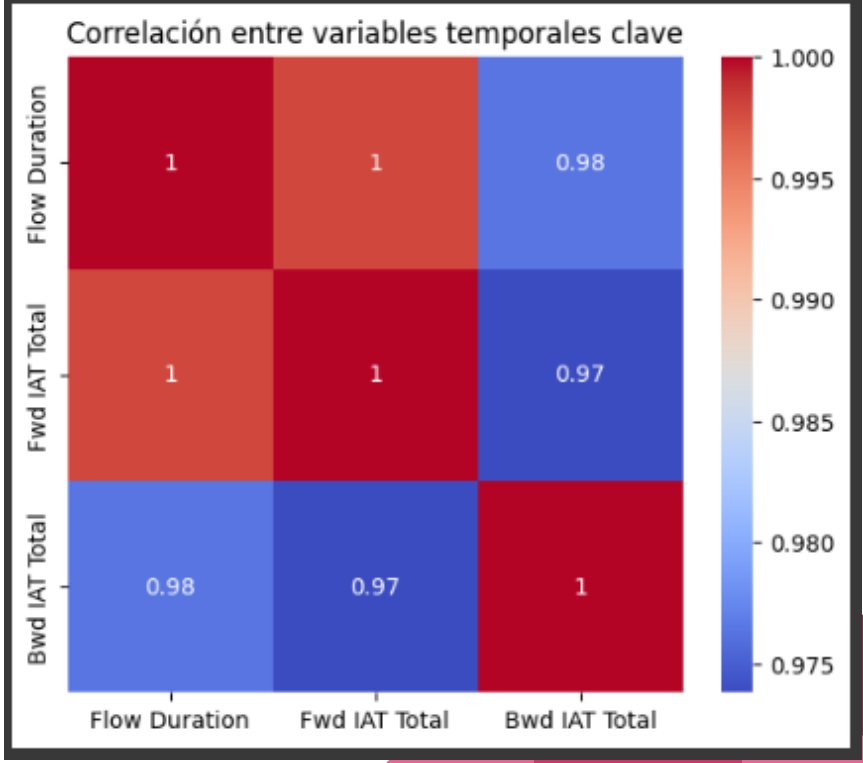


Label	
BENIGN	961727
FTP-Patator	7938
SSH-Patator	5897

¿Qué descubrieron al analizar los datos?

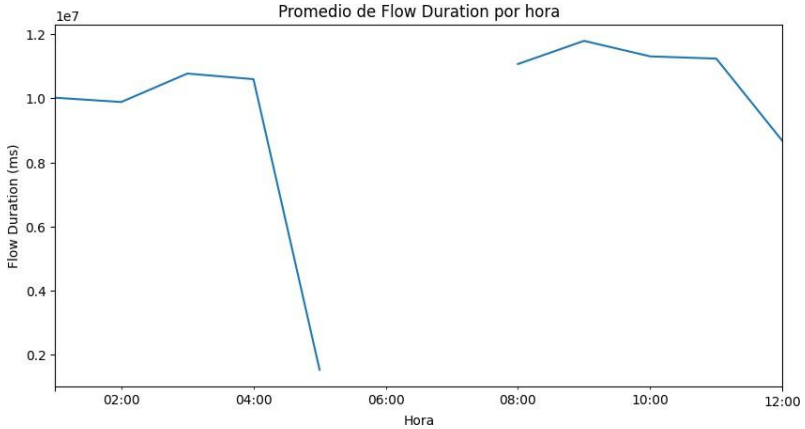
- Variables temporales como **Flow Duration**, **Fwd IAT Total** y **Bwd IAT Total** están altamente correlacionadas, mostrando que la duración y tiempos entre paquetes son factores clave.
- Variables con poca o ninguna variabilidad fueron entre tráfico benigno y malicioso. descartadas.
- El análisis estadístico y gráfico confirma diferencias claras

	Flow Duration	Fwd IAT Total	Bwd IAT Total
Flow Duration	1.000000	0.997916	0.976355
Fwd IAT Total	0.997916	1.000000	0.973811
Bwd IAT Total	0.976355	0.973811	1.000000



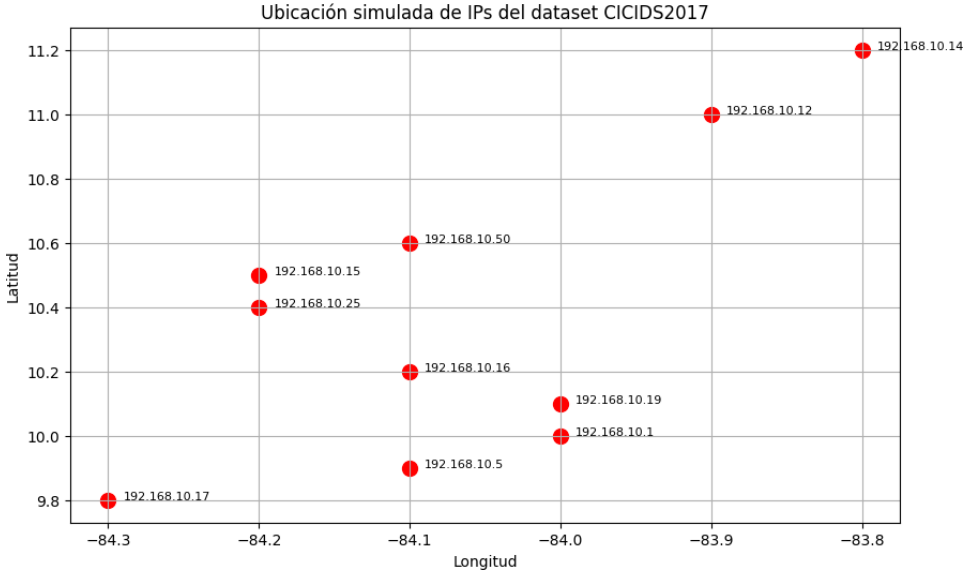
¿Qué reflejan los patrones de tendencia?

- El análisis de la variable **Flow Duration** a lo largo del tiempo muestra fluctuaciones significativas según la hora del día.
- Esto sugiere la existencia de picos de actividad relacionados con el comportamiento humano o ataques programados.
- El uso del campo **Timestamp** permitió hacer un análisis temporal agregando los datos por hora.



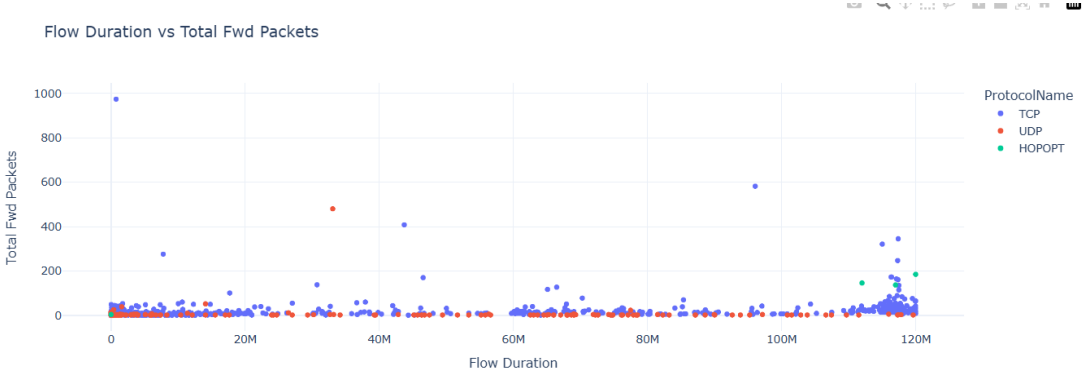
¿Cómo se ve afectado el comportamiento humano de movilidad en relación con la geografía?

- En este dataset, no contamos con datos geográficos, pero podemos usar atributos como **Protocolo**, **Puertos de destino** y **Horario** para inferir patrones de actividad humana en la red.
- Los gráficos muestran que el tráfico varía según protocolo y puerto, y que la actividad maliciosa tiene horarios específicos, reflejando comportamientos planificados o estacionales.



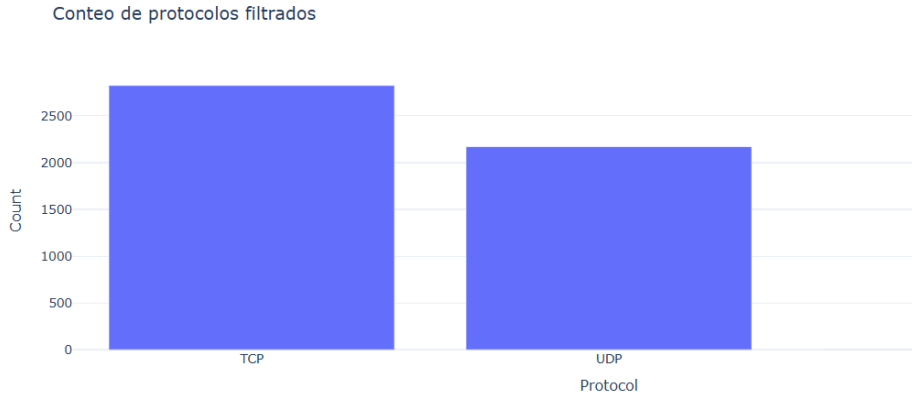
Hipótesis 1

Las conexiones maliciosas tienden a concentrarse en rangos específicos de duración de flujo y número de paquetes hacia adelante, diferenciándose del tráfico benigno que es más disperso.



Hipótesis 2

El uso de ciertos protocolos (como TCP o UDP) está fuertemente asociado a determinados tipos de ataques, y su distribución cambia cuando se filtra por etiquetas específicas.



Hipótesis 3

Las IPs maliciosas presentan patrones temporales claros de actividad que pueden visualizarse en líneas de tiempo y heatmaps horarios, indicando horarios preferidos para ataques o campañas.